ResearchGate

See discussions, stats, and author profiles for this publication at: https://www.researchgate.net/publication/233201173

The effectiveness of computer assisted pronunciation training for foreign language learning by children

Article *in* Computer Assisted Language Learning · December 2008

DOI: 10.1080/09588220802447651

CITATIONS
203

READS
4,321

4 authors:

Ambra Neri
imec
24 PUBLICATIONS **1,326** CITATIONS
SEE PROFILE

Ornella Mich
Fondazione Bruno Kessler
70 PUBLICATIONS **1,346** CITATIONS
SEE PROFILE

Matteo Gerosa
Fondazione Bruno Kessler
29 PUBLICATIONS **1,150** CITATIONS
SEE PROFILE

Diego Giuliani
Fondazione Bruno Kessler
105 PUBLICATIONS **2,123** CITATIONS
SEE PROFILE

All content following this page was uploaded by Ornella Mich on 12 February 2015.

The user has requested enhancement of the downloaded file.

Computer Assisted Language Learning
Vol. 21, No. 5, December 2008, 393–408

Routledge
Taylor & Francis Group

The effectiveness of computer assisted pronunciation training for foreign language learning by children

Ambra Neri, Ornella Mich*, Matteo Gerosa and Diego Giuliani

Center for Information Technology, Human Language Technologies Unit, Fondazione Bruno Kessler, Trento, Italy

(*Received 30 March 2007; final version received 22 May 2008*)

This study investigates whether a computer assisted pronunciation training (CAPT) system can help young learners improve word-level pronunciation skills in English as a foreign language at a level comparable to that achieved through traditional teacher-led training. The pronunciation improvement of a group of learners of 11 years of age receiving teacher-fronted instruction was compared to that of a group receiving computer assisted pronunciation training by means of a system including an automatic speech recognition component. Results show that 1) pronunciation quality of isolated words improved significantly for both groups of subjects, and 2) both groups significantly improved in pronunciation quality of words that were considered particularly difficult to pronounce and that were likely to have been unknown to them prior to the training. Training with a computer-assisted pronunciation training system with a simple automatic speech recognition component can thus lead to short-term improvements in pronunciation that are comparable to those achieved by means of more traditional, teacher-led pronunciation training.

Keywords: computer-assisted language learning (CALL); computer-assisted pronunciation training (CAPT); automatic speech recognition; children's speech; foreign language learning; pronunciation assessment

The importance of beginning to train a child's pronunciation skills in a second language (L2) at an early age has long been known to researchers and educators. In the past, the main reason was to capitalise on the advantages that children have over adults in learning this skill. It was believed that, within the window of a critical period extending from approximately two years of age to puberty, children could learn an L2 with little if any effort, in contrast to adults (Lennenberg, 1967). A large body of research has been conducted since (for an overview, see Birdsong, 1999). Several recent studies have shown that early exposure to an L2 can indeed lead to more accurate speech perception and production in that L2 than late exposure (see overview in Flege & MacKay, 2004).

However, nowadays we also know that the ability to learn an L2 is not lost abruptly, but rather declines linearly with age, and that pronunciation starts being affected by this loss early on (Khul, Williams, Lacerda, Stevens, & Lindblom, 1992; Polka & Werker, 1994). Children entering primary school have already acquired the bulk of their native

*Corresponding author. Email: mich@fbk.eu

ISSN 0958-8221 print/ISSN 1744-3210 online
© 2008 Taylor & Francis
DOI: 10.1080/09588220802447651
http://www.informaworld.com

language system and refined their phonetic-phonological system to such a degree that this development is already likely to hamper the acquisition of a new language (Flege, 1995).

In addition, today's policy on pronunciation training for children is obviously being shaped by the current social, economical, and political situation. If until a few decades ago speaking a foreign language (FL) seemed relatively dispensable and people could perfectly function as monolinguals in their work and everyday lives, being able to communicate in an FL has now become so crucial that the European Union has started taking measures to foster multilingualism in its member countries (BEC [Barcelona European Council], 2002). To comply with this policy, a number of member countries, including Italy, have recently made learning an FL compulsory from the first years of primary education onwards, with English being the most commonly taught language.

In summary, we now know that learning pronunciation in an L2 might not be as straightforward for children as was originally assumed. Furthermore, learning pronunciation in an L2 has become a fundamental requirement. Therefore, researchers and educators must devise optimal ways to provide pronunciation training for young learners.

What is the present situation, then, with respect to pronunciation training programmes for children? Partly as a result of current technological development, the use of computers to help children learn pronunciation skills in an L2 or FL has rapidly increased in the last decade. This is reflected by the presence on the market of commercial systems specifically designed for L2 pronunciation training for children, such as *English for kids* (Krajka, 2001) and the *Tell me more*/*Talk to me kids* series (TMM KIDS, 2001). The popularity of these systems is also motivated by the pedagogical requirements that computer-assisted pronunciation training (CAPT) can meet. CAPT systems can offer abundant, realistic, and contextualised spoken examples from different speakers by means of videos and recordings that learners can play as often as they wish. They can also provide opportunities for self-paced, autonomous practice, by inviting users to repeat utterances or to respond to certain prompts (see Neri, Cucchiarini, Strik, & Boves (2002), for an overview of how these requirements are implemented in current CAPT courseware).

This can be particularly beneficial in typical FL learning settings. In these instructional settings, exposure to oral examples in the target language is generally limited to the teacher's speech, and interaction with native speakers is often impossible. This lack of time available for contact with the language might be the most important reason for incomplete acquisition of the FL (Lightbown, 2000). Moreover, in these contexts, learning mainly takes place through the written medium, which might lead to stronger orthographic interference on pronunciation (Young-Scholten, 1997) than in contexts with more emphasis on oral communication.

Among CAPT systems, those incorporating automatic speech recognition (ASR) technology are attracting more and more interest (Bunnel, Yarrington, & Poliknoff, 2000; Chou, 2005; Eskenazi & Pelton, 2002; Giuliani, Mich & Nardon, 2003; Kawai & Tabain, 2000; Krajka, 2001; Sfakianaki, Roach, Vicsi, Csatari, Oster, & Kacic, 2001; TMM KIDS, 2001) because of a number of additional advantages that these systems can offer. Task-based speaking activities can be included, such as interactive speech-based games and role-plays with the computer, as in Auralog's *Tell me more*/*Talk to me kids* series (TMM KIDS, 2001) and in the system described in Bunnel et al.'s (2000) study. Such activities make learning a more realistic, rewarding, and fun experience (Purushotma, 2005; Wachowicz & Scott, 1999).

The most advanced systems incorporating ASR technology can also provide feedback at the sentence, word, or phoneme level. Automatic feedback can vary from rejecting poorly pronounced utterances and accepting 'good' ones to pinpointing specific errors

either in phonemic quality or sentence accent (e.g. Bunnel et al., 2000; Chou, 2005; Eskenazi & Pelton, 2002; TMM KIDS, 2001). This feedback can make the learner aware of problems in his or her pronunciation, which is the first necessary step to remedy those problems. Raising issues early on by means of automatic feedback might also prevent learners from developing wrong pronunciation habits that might eventually become fossilised (Eskenazi, 1999). As teachers have very little time to perform pronunciation evaluation and provide individual feedback in traditional language teaching contexts, the possibility to automate these tasks is considered one of the main advantages of ASR-based CAPT (Eshani & Knodt, 1998; Neri et al., 2002).

Not surprisingly, research into these systems has grown too. Some of the studies conducted have shown that children do seem to enjoy training pronunciation with ASR-based CALL and CAPT tools (e.g. Chou, 2005; Mich, Neri & Giuliani, 2005; Wallace, Russell, Brown, & Skilling, 1998). A considerable number of studies have also investigated the recognition and scoring accuracy of the ASR-based algorithms of CAPT systems for children (Eskenazi & Pelton, 2002; Gerosa & Giuliani, 2004; Hacker, Batliner, Steidl, Nöth, Niemann, & Cincarek, 2005; Steidl, Stemmer, Hacker, Nöth, & Niemann, 2003).

However, no empirical data have been collected, to our knowledge, on the actual pedagogical effectiveness of these systems for children. Research seems to be driven more by technological development rather than by the pedagogical needs of learners (Neri et al., 2002). As a result, systems with sophisticated features are built and sold, but we do not know whether the features and functionalities that they include will actually help learners to achieve better pronunciation skills. The need for assessing pedagogical effectiveness is a common, serious problem in CALL research in general (Chapelle, 1999; Chapelle, 2005; Felix, 2005) but it is even more acute in the case of ASR-based CAPT systems: recognising and evaluating non-native speech with current ASR technology still implies the risk of errors (Franco et al., 2000; Neri et al., 2002). Children's non-native speech represents an additional challenge because of the higher variability in its acoustic properties compared to adult speech (Gerosa & Giuliani, 2004; Hacker et al., 2005).

In order to gather evidence indicating whether CAPT systems for children can indeed offer valuable help towards the improvement of pronunciation skills, a CALL system with an ASR component was developed at the ITC-irst research institute (now FBK – Bruno Kessler Foundation) in Trento, Italy. The system, called PARLING (*PARla INGlese*, i.e. 'Speak English'), focussed on pronunciation quality at the word level. PARLING was tested by Italian children learning English within a real FL context. The purpose of the experiment was to establish if CAPT supported by ASR technology for children can lead to an improvement in pronunciation quality of isolated words, and if this improvement is comparable to that achieved in a traditional instructional setting in which the training is provided by a teacher. The remainder of this paper describes the operation of the system, the experiment conducted to test its effectiveness, and the results obtained.

The CAPT system considered: PARLING

Design

The design of PARLING was based on an analysis of relevant literature and of existing systems with similar purposes. For the latter analysis, *Tell me more, kids: The city* (TMM KIDS, 2001), was selected by language teachers and by researchers at ITC-irst. This system, which provides automatic feedback at the word and sentence level in four different modalities ranging from the presentation of oscillograms to animated characters, was deemed to meet most of the requirements set by these experts. Twenty-five 10-year-old

children were subsequently asked to use this system in a series of tests to study how they would interact with it, and to complete questionnaires on the system (Giuliani et al., 2003). The results of this preliminary study, together with the indications obtained from a pool of teachers and from an analysis of available literature, led to the development of PARLING.

More precisely, for the training focus, it was decided that PARLING should concentrate on pronunciation quality of isolated words in order to match the focus of the traditional training provided in regular classes. Moreover, during the collection of recordings used to fine-tune the ASR component, it was found that pronouncing English words in isolation already represents a challenging task for Italian beginner learners of English of the same age group as this study's subjects. Words were presented in their orthographic and audio form, in line with the recommendations in Giuliani et al. (2003). With respect to the feedback, it was decided to provide a simple accept/reject response (see below). This choice was motivated by results from the preliminary study: technically simple forms of feedback used in the tested system, such as digital waveform plotting (which are readily available in most signal processing software), and more sophisticated forms of feedback, such as animated characters changing according to the degree of pronunciation quality, were often found incomprehensible or uninformative by the children and the teachers.

The user interface

PARLING is a modular system. Each module is composed of a story, an adaptive word game based on the story, and a set of *active* words (see below). The system also includes a visual dictionary, a tool that allows children to create their own dictionary, and a simple help menu. Already from the start page of PARLING, users can access the stories, games, and dictionary, as well as tools to create a personal dictionary and, in the teacher's version, to build a new story (see Figure 1).

The stories are simplified versions of well-known children's stories. After choosing a story, the child can freely scan back and forth through its pages. Each time a page is loaded, its corresponding audio is played back. Each story comes with a different game meant to help the user memorise the words in that story. The game dynamically adapts its content to the user's personal work path.

Some words in these stories have hyperlinks so that when the user clicks on one of them, a window appears showing the meaning of the given word (see Figure 1). The user can optionally hear the pronunciation of the word as uttered by a British native speaker and try recording the word herself. The system analyses the recording in real time by means of ASR technology and responds with a message telling whether the word was pronounced correctly or not, and eventually prompting the child to repeat the incorrect utterance.

The dictionary in PARLING includes a tool with which the user can add new words. Children can type the new word of their choice, select a relevant image for it from an available database, and record the corresponding audio in their own voices. All operations performed by the users are logged. This way, a teacher can always monitor the children's work and progress.

The ASR component

The ASR component was based on context-independent Hidden Markov Models (HMMs) trained on read speech collected from native speakers of British English (aged

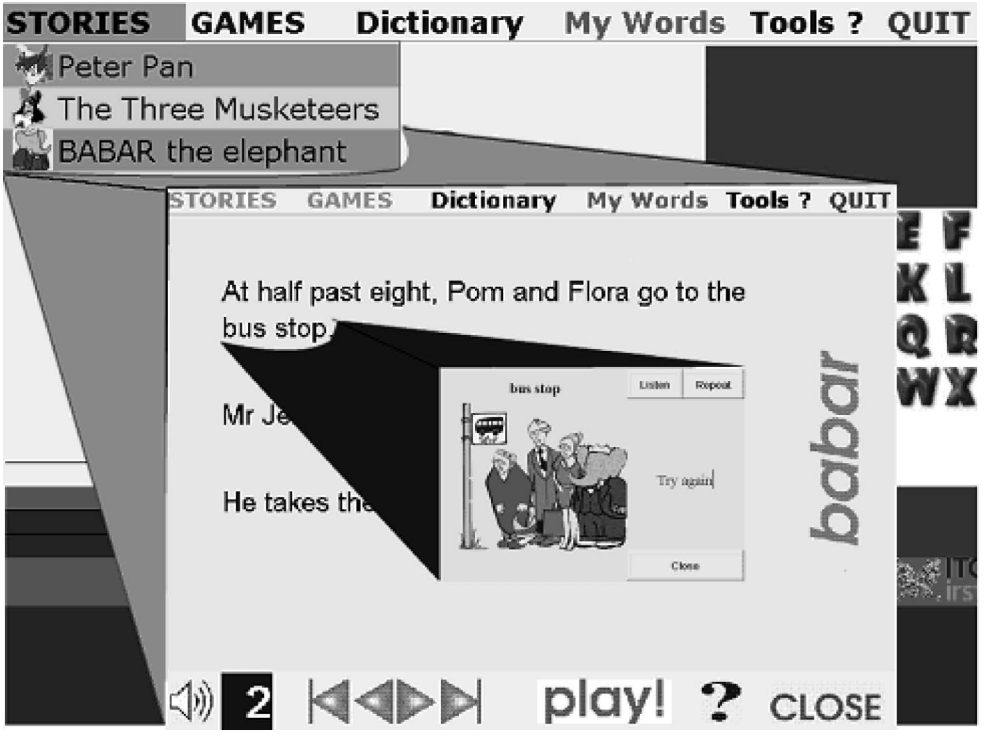

Figure 1. PARLING: the start page, a story page, and an active word in the story.

10 to 11) and adapted with read speech from Italian learners of English (aged 7 to 12) (Gerosa & Giuliani, 2004). This ASR component provided a simple accept/reject response for each input utterance. This was obtained in the following way. Each utterance was time-aligned with the sequence of HMMs corresponding to the phonemes of the canonical pronunciation of the uttered text. In doing this, pronunciation variants were taken into account. Phone recognition was then performed on the same utterance, adopting a simple phone-loop network with a heuristically determined phone-insertion penalty. Finally, the likelihood score achieved by the time alignment was compared to the likelihood achieved by the phone recognition. If the likelihood achieved with the forced time-alignment was higher than the likelihood achieved by phone recognition, the pronunciation was considered not too divergent from the expected standard pronunciation: in this case the ASR response was 'accept', otherwise it was 'reject' (see Gerosa & Giuliani, 2004, for a more detailed description of the ASR system used). These responses were, respectively, 'Well done!' and 'Try again' on the user interface.

Method

To measure and compare possible improvements on pronunciation quality of words after four weeks of traditional training and of training with PARLING, two groups of Italian children were studied before and after the training. The control group received instruction in the form of traditional, teacher-led classes. The experimental group worked with PARLING during individual sessions.

Subjects

The 28 subjects were all 11-year-old Italian native speakers attending the same public school and sharing the same curriculum. They studied in two different groups, but they were attending the same type of classes and had the same English teacher. At the time of the experiment, they all had had four years of English FL classes. Group C, i.e. the control group, was composed of 15 children, while group E, i.e. the experimental group, included 13 children.

Training procedure

Group C participated in four teacher-led (British) English FL class sessions of 60 minutes each. During each session, the teacher read an excerpt from a simplified, English version of the Grimms' children's story *Hansel and Gretel*. This story was chosen because Italian children of 11 are generally familiar with it, which could thus help them to more easily understand the corresponding English version which was included in the training. Participants in this group were provided with a printed version of the story. The teacher also discussed some words found in the story with the children, explaining their meaning and providing his rendition of the correct pronunciation. He regularly prompted the children to repeat words aloud, mostly as a group. At the end of each training session, each child also completed a printed word game based on words extracted from the excerpt of the story that had been read in that session.

The children in group E had four individual CAPT training sessions in the school's language lab, each lasting 30 minutes, during which they worked with PARLING. The limited amount of time for these sessions was due to the limited number of computers available in the language lab: children had to take turns. In order for the training to be comparable to that received by the subjects in group C, the experimental group worked with a modified version of PARLING. This version did not include the dictionary tool and only contained one story – the same one studied by the control group – which was divided into four parts, one for each training session. During each session, the children listened to the relevant part of the story while reading it on the screen. They listened to and repeated some of the words presented in that session's excerpt. These active words ($n = 41$) had hyperlinks that allowed the children to listen to the word's pronunciation as often as they wished. A minimum of one recording for each word was mandatory for the children to continue the session. The word had to be repeated until a positive response was received or until a maximum of four negative attempts was reached. The system would only move to the next page after a child had repeated all the active words of a page. At the end of each story excerpt, children also played a word game that only included words presented in the story excerpt of that session. For this game, children had to pronounce and record the words proposed by the game. If the spoken utterance was rejected by the ASR module, the child had to repeat the word at least one more time.

In this way, the only difference was that the training was provided by a teacher in the case of group C, and by a computer in the case of group E.

Testing procedure

In order to be able to evaluate and compare the participants' possible improvements in pronunciation quality at the word level, we asked all children to read and record a set of

28 isolated words (see Appendix) before and after the training. These were subsequently scored by three experts. Read speech was chosen as the elicitation material to allow for comparisons across subjects. The words were taken from the simplified version of the story with which the children were presented during the training. The words were chosen so as to cover the most frequent British English phonemes. These words varied with respect to length, articulatory difficulty, and lexical frequency. The children's English FL teacher was therefore asked to indicate which of the 28 words might have been more difficult for the children to pronounce (e.g. because of likely negative orthographic interference, or consonant clusters that are unusual or difficult to articulate for native speakers of Italian), and which were unlikely to have been known to the children before the training. Based on these indications, a matrix was obtained of easy ($n = 21$), difficult ($n = 7$), known ($n = 21$), and unknown words ($n = 7$). The higher number of easy and known words was a consequence of the text stimulus chosen, i.e. a well-known, simplified children's story. This disparity would nevertheless ensure the ecological validity of the experiment, as it realistically reflected the type of stimuli the children are exposed to in their regular English FL classes.

For the recording sessions, a dedicated tool was used that presented one word at a time on the screen, prompted the child to read it aloud, and recorded it. If the child felt that he or she had not pronounced the word correctly, he or she was allowed to repeat it and record it as many times as needed. The recordings were made with head-mounted microphones and the speech was sampled at 16 kHz.

Rating procedure

The recordings were scored independently by three native speakers of British English who were working in Italy as English FL teachers, based on the procedure described in Cucchiarini, Strik, and Boves (2000) and Neri, Cucchiarini, and Strik (2008). Each rater was asked to provide a score of pronunciation quality of each utterance on a 10-point scale.

Two different types of ratings were requested: a rating for each word, and a rating for each speaker. For the former, each recording corresponding to one word was presented in a separate audio file and had to be scored individually. For the latter, all words recorded by each participant were concatenated in random order in a larger audio file, for which one score was elicited, so that one score would be available per speaker. An additional scoring round was also organised using 32 duplicate recordings to allow the calculation of intra-rater reliability. The duplicates were selected semi-randomly from the whole set of recordings. More precisely, for the single-word scores, one recording per subject was randomly selected out of that subject's single-word recordings. In addition, four recordings were randomly selected out of the whole set of speaker (concatenated) recordings.

Raters were allowed to complete the task in several sessions, so as to avoid possible fatigue effects in the scores. The duplicates were presented to the raters two weeks after they had completed the rating sessions. To help the raters familiarise with the scoring scale, examples of spoken words of 'very poor' pronunciation quality were provided at the beginning of a rating session, together with examples of words produced by a native speaker of English, i.e. words of very good quality pronunciation. The single-word audio files were presented in 28 blocks, with each block containing the same word uttered by all subjects at both testing conditions. The audio files in each block were presented in random order. In total, each rater assigned 1656 scores (see Table 1).

Results

Reliability of ratings

The raters' scores were first analysed to determine inter- and intra-rater reliability.

The computation of inter-rater reliability was based on 1568 scores from each rater (28 participants × 2 testing conditions × 28 words). A Cronbach's alpha coefficient of .872 was obtained, which can be considered satisfactory. Intra-rater reliability was calculated on the basis of 32 (16 pre-test files + 16 post-test files) scores for each rater. The intra-rater reliability coefficients for the three raters ranged from .757 to .859 (see Table 2), indicating, again, high reliability.

However, since the distributions of the scores assigned by each rater were rather different (see Figure 2), we decided to standardise the scores of each rater separately, based on the means and standard deviations of the scores provided by that rater alone. We then averaged *z*-scores to obtain one score for each word and for each speaker, as indicated in Cucchiarini et al. (2000). The standardised scores were used for the remainder of the analyses.

Pronunciation quality of words

Before analysing possible improvements in the two groups, we looked at how the speaker scores related to the single-word scores, by comparing the former with the mean of the single-word scores for each speaker. In other words, in one case we had one score assigned by three raters to one participant, and in the other case we had one score per participant that was obtained by averaging all single-word scores for that participant. The check was carried out because the two types of scores might have differed: it might have been possible that, for instance, few seriously mispronounced words that happened to be located at the end of a concatenated audio file affected more severely the speaker scores than the mean of the single-word scores. We found a strong, positive correlation between the two types of scores ($r = .884$, $p < .01$) (see Figure 3).

Table 1. Audio files scored by each rater.

	Single words	Total	Concatenated words (speakers)	Total	Grand total
Pre-test	28 subjects × 28 recordings	784	28 subjects × 1 recording	28	812
Post-test	28 subjects × 28 recordings	784	28 subjects × 1 recording	28	812
Duplicates					
(pre-test)	14 recordings	28	2 recordings	4	32
(post-test)	14 recordings		2 recordings		
		1596		60	1656

Table 2. Reliability coefficients (Cronbach's alpha).

	Intra-rater		
Inter-rater	Rater 1	Rater 2	Rater 3
.872	.859	.764	.757

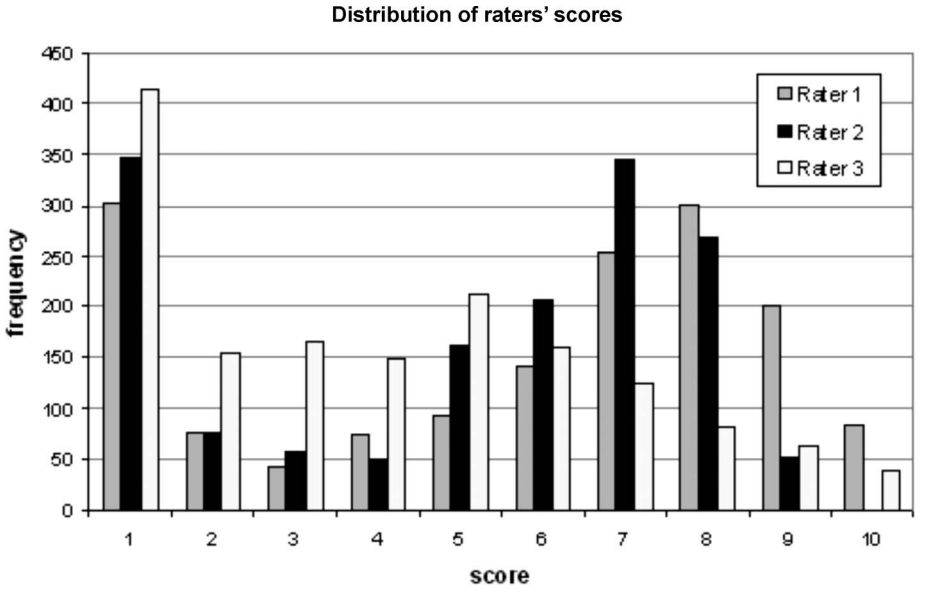

Figure 2. Distribution of pronunciation scores assigned to single words by the three raters.

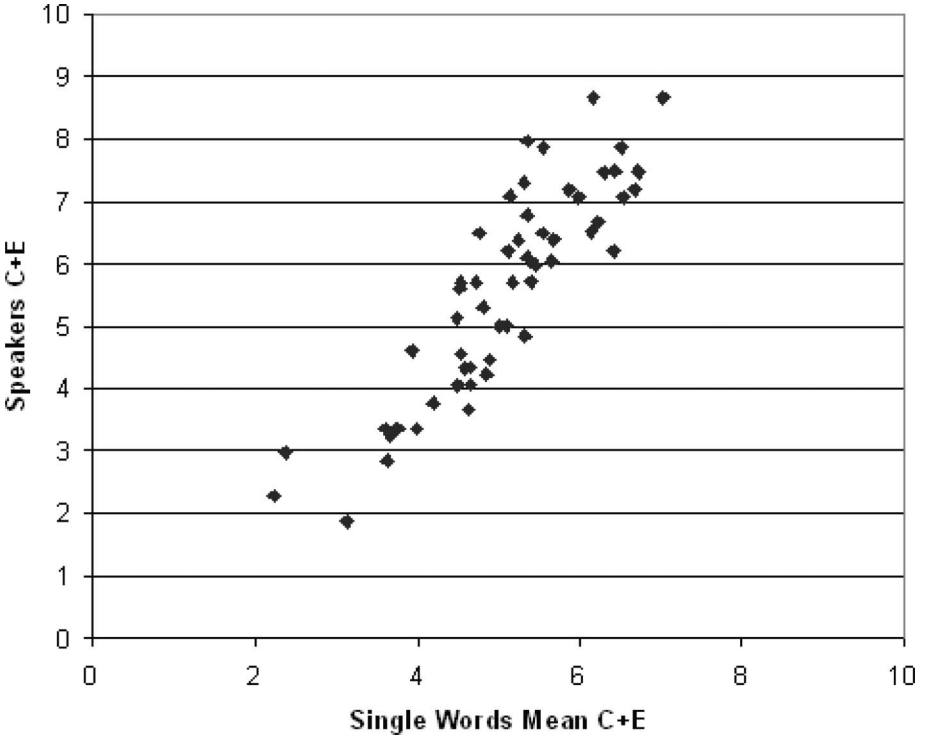

Figure 3. Correlation between single-word and speaker scores.

For group C, the results were: $r = .904$, $p < .01$, for group E: $r = .850$, $p < .01$. We therefore assumed that the speaker scores were a good reflection of overall pronunciation quality of the words selected and could thus be used for the remainder of the analyses.

We then carried out a *t*-test on the participants' speaker scores on pronunciation quality *prior* to the training. Results ($t = .321$, $p = .754$) indicate that pronunciation quality at word level in the two groups was not significantly different at pre-test. The overall mean scores at pre-test for group C ($M = 4.57$, $SD = 1.55$) and for group E ($M = 4.41$, $SD = 1.12$) were similar (see Table 3).

We thus proceeded to analyse the participants' speaker scores to assess possible improvements in pronunciation quality of isolated words after the training, and possible differences between control and experimental groups in this respect. An ANOVA with repeated measures[1] with test time (levels: pre-test, post-test) as the within-subjects factor and training group (levels: C, E) as the between-subjects factor indicated a main effect for test time, with $F(1,26) = 78.818$, $p < .05$. The overall mean score at post-test ($M = 6.67$, $SD = 1.26$) was significantly higher than at pre-test ($M = 4.49$, $SD = 1.34$). No significant effect was found for training group, $F(1,26) = .610$, $p = .442$, nor was there a significant test × training interaction, with $F(1,26) = .548$, $p = .446$. These results indicate that both groups improved in pronunciation quality, and that their improvements were comparable (see also Figure 4).

Pronunciation quality of specific types of words

In order to gain more insight into the effectiveness of the training provided, we also examined the scores by looking at specific types of words. Recall that we had a matrix of easy, difficult, known, and unknown words (see testing procedure). Of this matrix, we retained the easy/known words ($n = 19$) and the difficult/unknown words ($n = 5$) (see Appendix for the complete word sets). The rationale behind this analysis was that the impact of the training provided for this experiment might more clearly emerge in the words that were unknown to the children prior to the training and that were particularly difficult to pronounce. At pre-test, these words had an average rating of 3.01 ($SD = 2.47$) and 2.99 ($SD = 2.29$) for group C and group E, respectively, against 5.12 ($SD = 2.86$) and 4.88 ($SD = 2.90$) of known/easy words.

We thus submitted these scores to an ANOVA with repeated measures involving test time (levels: pre-test, post-test) and word type (levels: easy/known, difficult/unknown) as the within-subjects factors, and training group (levels: C, E) as the between-subjects factor. This analysis revealed a significant effect for the test, with $F(1,26) = 144.729$, $p < .01$. A significant, main effect was also found for word type ($F(1,26) = 57.531$, $p < .01$) with the mean scores of the easy/known words ($M = 5.44$, $SD = 0.93$) being significantly higher

Table 3. Pre-test and post-test speaker scores for the two groups.

	Group	Mean	SD	N
Pre-test	**C**	4.57	1.55	15
	E	4.41	1.12	13
	Overall	4.49	1.34	
Post-test	**C**	6.91	1.37	15
	E	6.39	1.12	13
	Overall	6.67	1.26	

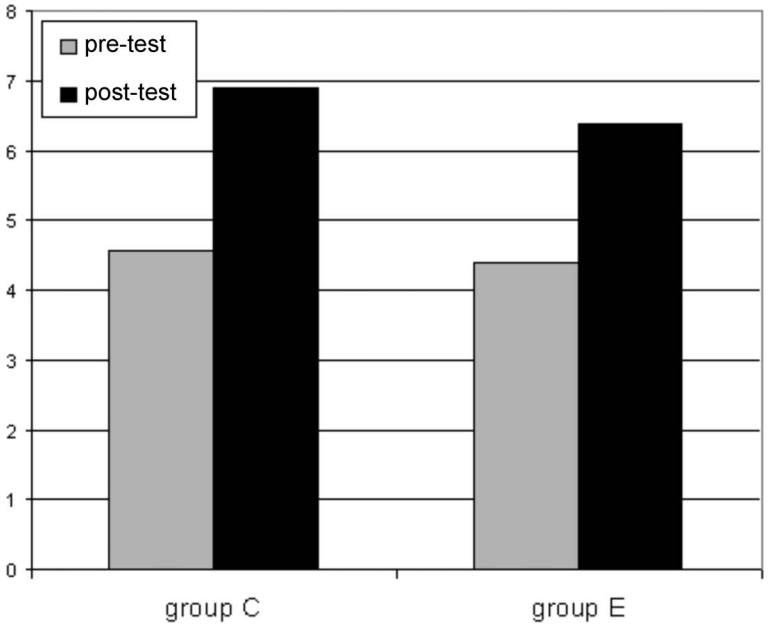

Figure 4. Speaker scores group means for C and E before and after the training.

than those of the difficult/unknown words (M = 4.32, SD = 1.01). An interaction between test time and word type was also found, with F(1,26) = 60.080, $p < .01$, with the pre-test mean of the difficult/unknown words (M = 3.06, SD = 1.10) being significantly lower than the mean of the easy/known words (M = 5.11, SD = 1.08), while at post-test, the means of difficult/unknown words and easy/known words are not significantly different (M = 5.59, SD = 0.93, M = 5.74, SD = .79, respectively).

In other words, the pronunciation quality of difficult/unknown words improved significantly after the training, whereas that of easy/known words did not. Since no significant effect for test × word × training was found (F(1,26) = 3.078, $p = .091$), we can conclude that the improvements of the two groups were not significantly different for the two types of words (see Figure 5).

Discussion and conclusions

The use of CAPT and CALL systems to support pronunciation training in a second/foreign language is becoming more and more widespread. However, very little is known about the actual pedagogical effectiveness of these systems, especially when young learners are considered. The purpose of this study was to evaluate the pedagogical effectiveness of a CAPT system with ASR technology for children. To this aim, we first measured the improvement achieved by 11-year-old learners of English in pronunciation quality of isolated words. We then compared this improvement to that achieved by a similar group of learners in a traditional instructional setting in which the training was provided by a teacher.

The rationale behind this analysis was that the learning gain achieved within the traditional instructional setting can be considered the benchmark against which to gauge the pedagogical effectiveness of the computer-based training. Obviously, it is unrealistic to

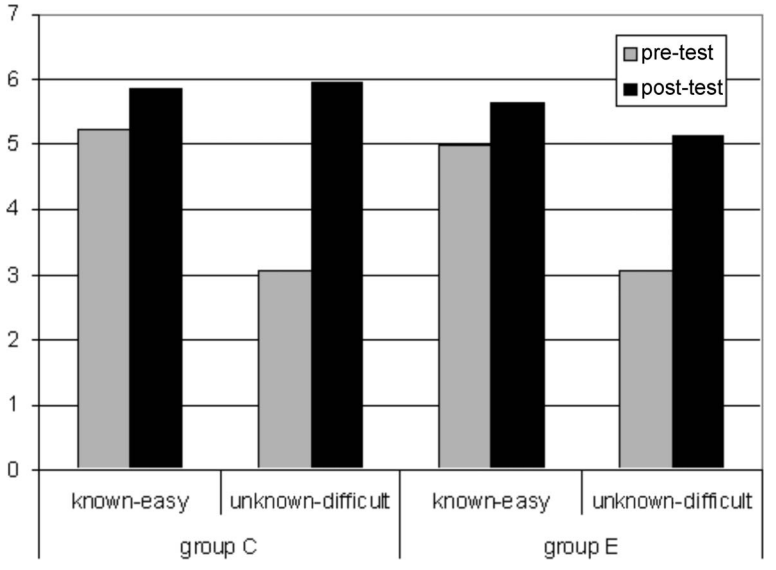

Figure 5. Group means for C and E before and after the training for known/easy words and for unknown/difficult words. The values were obtained by averaging the corresponding single-word *z*-scores.

expect that CAPT or CALL systems can perform all the tasks that a teacher can perform with the same effectiveness. However, the use of these systems can only be justified if it demonstrably leads to benefits that are comparable or at least close to those achieved when training is provided by a teacher.

The results of the analysis on speaker scores show that the children who trained with PARLING were able to improve pronunciation quality of isolated words. This improvement was found to be comparable to that achieved by the children who received teacher-led instruction with the same focus. This finding is all the more positive considering that the children training with PARLING could only train for 30 minutes per session, while the children in the control group had 60-minute sessions, and that the automatic feedback provided on pronunciation in PARLING was a simple reject/accept response. The analysis carried out on different types of words further indicates that the children in the two groups also made comparable improvements in pronunciation quality of words which they did not know before the training, and which were particularly difficult to pronounce. These positive results might be explained by the fact that the children using PARLING enjoyed the computer's 'undivided attention' for all 30 minutes of training, while the children training with the teacher could seldom practice and receive feedback individually during the 60-minute lesson.

A few limitations of this study should nevertheless be mentioned. First of all, pronunciation quality of isolated words is only one aspect of pronunciation skills. In this study, this aspect was chosen because it was indicated as the main focus of FL training for Italian children in the age group considered. However, being able to correctly imitate isolated words does not imply being able to correctly pronounce sentences in spontaneous connected speech, where factors such as co-articulation and sentence accent play an important role. Further research in CAPT for children should address pronunciation aspects beyond the word level. It would also be interesting to differentiate between the

segmental and the suprasegmental levels. Obviously, the more factors that are addressed in the training, the more complex it becomes to ascertain their relative impacts on the results.

Another aspect that should be studied further is the impact of the type of feedback provided on the learning gain. In this study, a simple form of feedback was chosen due to problems previously evidenced with other forms of feedback. However, one can imagine that a more composite type of feedback detailing the main problems in an utterance, and possibly directing the user to further practice on those specific problems, might be more helpful to correct problems globally. In order to establish that, different types of feedback should be provided to different learners and the relative improvements measured should be compared.

Finally, it should be pointed out that the results obtained in this study only pertain to the short-term effects of the training and that the sample size considered is relatively small. Additional studies with delayed post-tests might add robustness and generalisability to our results.

Despite the limitations above, the findings from this study have provided the first empirical evidence that CAPT training can be effective in helping children to improve pronunciation skills. These results have pedagogical implications: CAPT systems could be used to integrate traditional instruction, for instance to alleviate typical problems due to time constraints or to particularly unfavourable teacher/student ratios. In this way, children could benefit from more intensive exposure to oral examples in the FL, and from more intensive individualised practice and feedback on pronunciation in the FL. This would free up time for the teacher, which could be employed to provide individual guidance on how to remedy specific pronunciation problems – something computers are not yet capable of doing in a reliable way. Another opportunity would be to provide (customised) CAPT for children who are lagging behind, thus offering them an additional, engaging, and more private form of training.

Acknowledgements

A reduced version of this paper was presented at the 12th International CALL Research Conference, 2006, Antwerp, Belgium. Ambra Neri's contribution to the research reported in this paper was made possible by a *Frye Stipendium* granted by the Radboud University Nijmegen, the Netherlands.

We would like to thank Sergio Amadori, for his constant cooperation throughout this study, the primary school *Paride Lodron* of Villa Lagarina (Italy), and above all the children who took part in the experiment described in this paper, for their dedication and enthusiasm. We would also like to thank two anonymous reviewers for their comments on an earlier version of this paper.

Note

1. All assumptions (normality, homogeneity of variance and covariances) of the ANOVAs described in this study were met.

Notes on contributors

Ambra Neri obtained a PhD from the Department of Linguistics of the Radboud University Nijmegen, the Netherlands, with a thesis on the pedagogical effectiveness of ASR-based Computer Assisted Pronunciation Training for adults.

Ornella Mich received a degree in Electronic Engineering from the University of Padova, Italy. She is currently pursuing a PhD course in Computer Science. Her main research interests focus on interaction design and children, and computer accessibility.

Matteo Gerosa received a degree in Physics and a PhD in Computer Science from University of Trento, Italy. His research focuses on automatic speech recognition, speech analysis, speech understanding and acoustical modeling.

Diego Giuliani received a Laurea degree in Computer Science from the University of Milan, Italy. Be has been doing research in robust speech recognition, speaker adaptation/normalization, acoustic modeling and children's speech recognition since 1988.

References

BEC (Barcelona European Council) (2002). *Presidency Conclusions. Barcelona European Council 15 and 16 March 2002, part 1, 43.1*. Retrieved February 27, 2008, from http://www.consilium.europa.eu/ueDocs/cms_Data/docs/pressData/en/ec/71025.pdf.

Birdsong, D. (1999). *Second language acquisition and the critical period hypothesis*. Mahwah, NJ: Lawrence Erlbaum Associates.

Bunnel, H.T., Yarrington, D.M., & Poliknoff, J.B. (2000). STAR: Articulation training for young children. Proceedings of ICSLP2000, 6th International Conference of Spoken Language Processing, Beijing, China, 2000. vol. 4 (pp. 85–88) China; China Military Friendship Publ.

Chapelle, C.A. (1999). Research questions for a CALL research agenda: A reply to Rafael Salaberry. *Language Learning and Technology*, *3*(1), 108–113. Retrieved February 27, 2007, from http://llt.msu.edu/vol3num1/comment/reply.html.

Chapelle, C.A. (2005). Computer assisted language learning. In E. Hinkle (Ed.), *Handbook of research in second language learning and teaching* (pp. 743–756). Mahwah, NJ: Laurence Erlbaum Associates.

Chou, F. (2005). Ya-Ya language box. A portable device for English pronunciation training with speech recognition technologies. Paper presented at proceedings of Interspeech-2005, Lisbon, Portugal (pp. 169–172). Retrieved from http://www.isca-speech.org/archive/.

Cucchiarini, C., Strik, H., & Boves, L. (2000). Quantitative assessment of second language learners' fluency by means of automatic speech recognition technology. *Journal of the Acoustical Society of America*, *107*(2), 989–999.

Eshani, F., & Knodt, E. (1998). Speech technology in computer-aided language learning: Strengths and limitation of a new call paradigm. *Language Learning & Technology*, *2*(1), 45–60. Retrieved February 27, 2007, from http://llt.msu.edu/vol2num1/article3/index.html.

Eskenazi, M. (1999). Using automatic speech processing for foreign language pronunciation tutoring: Some issues and a prototype. *Language Learning & Technology*, *2*(2), 62–76. Retrieved February 27, 2007, from http://llt.msu.edu/vol2num2/article3/index.html.

Eskenazi, M., & Pelton, G. (2002). Pinpointing pronunciation errors in children's speech: examining the role of the speech recognizer. Paper presented at Pronunciation Modeling and Lexicon Adaptation for Spoken Language Technology Workshop, Estes Park, Colorado (pp. 48–52).

Felix, U. (2005). Analysis of recent CALL effectiveness research. Towards a common agenda. *Computer Assisted Language Learning*, *18*(1–2), 1–32.

Flege, J.E. (1995). Second language speech learning: Theory, findings, and problems. In W. Strange (Ed.), *Speech perception and linguistic experience: Issues in cross-linguistic research* (pp. 233–277). Timonium, MD: York Press, Inc.

Flege, J.E., & MacKay, I.R.A. (2004). Perceiving vowels in a second language. *Studies in Second Language Acquisition*, *26*(1), 1–34.

Franco, H., Abrash, V., Precoda, K., Bratt, H., Rao, R., & Butzberger, J. (2000). The SRI EduSpeak system: Recognition and pronunciation scoring for language learning. Proceedings of InSTIL 2000, Dundee, Scotland (pp. 123–128). Dundee: Dundee University Press.

Gerosa, M., & Giuliani, D. (2004). Preliminary investigations in automatic recognition of English sentences uttered by Italian children. In R. Delmonte, P. Delcloque & S. Tonelli (Eds.), Proceedings of NLP and Speech Technologies in Advanced Language Learning Systems Symposium, Venice, Italy (pp. 9–12). Padova: Unipress.

Giuliani, D., Mich, O., & Nardon, M. (2003). A study on the use of a voice interactive system for teaching English to Italian children. Proceedings of the 3rd IEEE International Conference on Advanced Learning Technologies, Athens, Greece, (pp 376–377). Los Alamitos, CA: IEEE Computer Society Press.

Hacker, C., Batliner, A., Steidl, S., Nöth, E., Niemann, H., & Cincarek, T. (2005). Assessment of non-native children's pronunciation: Human marking and automatic scoring. In G. Kokkinakis, N. Fakotakis & E. Dermatas (Eds.), Proceedings of the 10th International Conference on SPEECH and COMPUTER (SPECOM 2005), Patras, Greece (pp. 123–126). Vol. 1. Moscow: Moskow State Linguistics University.

Kawai, G., & Tabain, M. (2000). Automatically detecting mispronunciations in the speech of hearing-impaired children. Proceedings of InSTIL 2000, Dundee, Scotland (pp 39–48). Dundee: Dundee University Press.

Krajka, J. (2001). English for Kids. *CALICO Software Review*, Retrieved January 19, 2007 from http://calico.org/CALICO_Review/review/englishkids00.htm.

Kuhl, P.K., Williams, K.A., Lacerda, F., Stevens, K.N., & Lindblom, B. (1992). Language experience alters phonetic perception in infants by 6 months of age. *Science*, *255*(5044), 606–608.

Lennenberg, E.H. (1967). *Biological Foundations of Language*. New York: Wiley and Sons.

Lightbown, P.M. (2000). Classroom SLA research and second language teaching, Anniversary article. *Applied Linguistics*, *21*(4), 431–462.

Mich, O., Neri, A., & Giuliani, D. (2005, September). Testing a system that supports children in learning English as a FL: Implications for a more effective design. Paper presented at the INTERACT 2005 Conference: WS7 Child Computer Interaction: Methodological Research, Rome, Italy.

Neri, A., Cucchiarini, C., & Strik, H. (2008). The effectiveness of computer-based corrective feedback for improving segmental quality in L2-Dutch, *ReCALL*, *20*(2), 225–243.

Neri, A., Cucchiarini, C., Strik, H., & Boves, L. (2002). The pedagogy-technology interface in Computer Assisted Pronunciation Training. *Computer Assisted Language Learning*, *15*(5), 441–467.

Polka, L., & Werker, J.F. (1994). Developmental changes in the perception of nonnative vowel contrasts. *Journal of Experimental Psychology: Human Perception and Performance*, *20*(2), 421–435.

Purushotma, R. (2005). Commentary: You're not studying, you're just ... *Language Learning and Technology*, *9*(1), 80–96. Retrieved February 27, 2007, from http://llt.msu.edu/vol9num1/purushotma/default.html.

Sfakianaki, A., Roach, P., Vicsi, K., Csatari, F., Oster, A.-M., Kacic, Z., & Barczikay, P (2001). SPECO: Computer-based Phonetic Training for Children. In J.A. Maidment & E. Estebas-Vilaplana (Eds.), Proceedings of the Phonetics Teaching & Learning Conference, London, 2001 (pp. 43–46). London: Department of Phonetics and Linguistics, UCL.

Steidl, S., Stemmer, G., Hacker, C., Nöth, E., & Niemann, H. (2003). Improving children's speech recognition by HMM interpolation with an adults' speech recognizer. In B. Michaelis & G. Krell (Eds.), Pattern Recognition, Proceedings of the 25th DAGM Symposium, Magdeburg, Germany, 2003. Lecture Notes in Computer Science, vol. 2781 (pp. 600–607). Berlin: Springer-Verlag.

TMM KIDS (2001). *Tell me more, KIDS, 8–10 anni. Impara l'inglese divertendoti! Kalikò's City, Auralog*. Milan: Opera (TM) Multimedia; Montigny-le-Bretonneux: Aurolog.

Wachowicz, K.A., & Scott, B. (1999). Software that listens: It's not a question of whether, it's a question of how. *CALICO Journal*, *16*(3), 253–276.

Wallace, J.L., Russell, M., Brown, C., & Skilling, A. (1998). Applications of speech recognition in primary school classroom. Proceedings of the ESCA Workshop on Speech Technology in Language Learning (STiLL), Marholmen, Sweden (pp. 21–24). Stockholm: ESCA.

Young-Scholten, M. (1997). Second-language syllable simplification: Deviant development or deviant input? In J. Leather & J. Allan (Eds.), New Sounds 97: Proceedings of the Third Symposium on the acquisition of Second-Language Speech, Klagenfurt, 1997 (pp. 351–360). Klagenfurt: University of Klagenfurt Press.

Appendix

Words used as speech stimuli (n = 28):

Away, birds, biscuits, breadcrumbs, buy, cage, cold, dinner, door, father, fire, food, good, home, house, hungry, idea, jumps, leave, locked, morning, pebbles, pushes, treasure, sweets, stepmother, witch, woodcutter

Known words (n = 21):

Birds, biscuits, buy, cage, cold, dinner, door, father, fire, food, good, home, house, hungry, idea, jumps, leave, morning, sweets, treasure, witch

Unknown words (n = 7):

Away, breadcrumbs, locked, pebbles, pushes, stepmother, woodcutter

Easy words (n = 21):

Away, birds, biscuits, buy, cage, cold, dinner, door, father, fire, food, good, home, house, hungry, jumps, leave, morning, pebbles, sweets, witch

Difficult words (n = 7):

Breadcrumbs, idea, locked, pushes, stepmother, treasure, woodcutter

Copyright of Computer Assisted Language Learning is the property of Routledge and its content may not be copied or emailed to multiple sites or posted to a listserv without the copyright holder's express written permission. However, users may print, download, or email articles for individual use.

View publication stats